Balkans JETSS (2025) 2: 123-137
DOI: https://doi.org/10.31410/Balkans.JETSS.2025.8.2.123-137

Original Scientific Article

Enhancing Hotel Management Through Predictive AI Models for Customer Lifetime Value (CLV)

Milena Nikolić[1]
Žarko Rađenović[2]
Marina Marjanović[3]

Received: October 10, 2025 / Revised: December 28, 2025 /
Accepted: December 29, 2025 / Published: December 30, 2025

Abstract: *This paper introduces a hybrid artificial intelligence (AI) framework for predicting Customer Lifetime Value (CLV) in hotel management. CLV represents the long-term financial contribution of guests and provides information for resource allocation, customer retention, and profitability. Traditional models rely on structured reservation records, often overlooking emotional insights in online reviews. To address this concern, the study combines structured booking attributes with unstructured guest reviews, using RoBERTa embeddings for textual data and XGBoost for numerical features. The proposed multimodal model achieves 89% accuracy, 87% precision, and 86% recall, outperforming single-source approaches utilizing bookings (63%) or guest reviews only (72%). SHAP-based interpretability reveals that review topics, including cleanliness, staff professionalism, and service quality, directly influence CLV, alongside structured features such as repeat bookings and special requests. The findings highlight the potential of predictive AI to enhance hotel management by identifying valuable customers early, supporting personalized services, and optimizing strategic decision-making.*

Keywords: *Hotel management, Customer Lifetime Value, Predictive AI, Multimodal Learning, RoBERTa, XGBoost*

JEL Classification: C45 · L83 · M31

✉ milena.nikolic@akademijanis.edu.rs
[1] Academy of Applied Technical and Preschool Studies, 18000, Niš, Serbia
[2] University of Niš, Innovation Center, Univerzitetski trg 2, 18000, Niš, Serbia
[3] Singidunum University, Danijelova 32, 11000, Belgrade, Serbia

1. INTRODUCTION

The modern hospitality industry continuously relies on data-driven decision-making to remain competitive and sustainable. The complexity of client journeys has increased with the rapid expansion of online booking platforms and review systems, which are driven by factors such as trust, loyalty, and reputation. Customer Lifetime Value (CLV) has become a central metric for strategic management, as it estimates the financial contribution of a guest throughout their relationship with a hotel. Early identification of high-value customers is crucial for optimizing pricing strategies, improving personalization, and reinforcing loyalty initiatives. Research shows that CLV frameworks provide essential guidance for modern hospitality applications, and consumer experiences directly participate in building long-term trust (Webb et al., 2022). This study extends prior research by conducting a deeper investigation into the role of CLV in the hotel industry.

CLV can be expressed either as a continuous monetary estimate or through categories that are easier to apply in practice. Table 1 illustrates three example CLV category groups, along with their associated revenue levels, booking patterns, and retention probabilities.

Table 1. Example Customer Lifetime Value categories and metrics

CLV Category	Average revenue per guest (USD)	Repeat booking rate	Retention probability	Illustrative value (USD)
Low	200	5%	0.25	< 500
Medium	600	20%	0.55	500 – 1500
High	1500	45%	0.80	> 1500

Source: Own processing

Artificial intelligence has further transformed this landscape by introducing predictive models that strengthen managerial strategies. Kabir et al. (2024) indicate that AI-driven forecasting analytics significantly improve decision quality across hotel operations, while Henriques and Pereira (2024) further confirm that embedding predictive systems into business practices enhances financial performance. These findings show that AI-driven models can forecast churn, loyalty, and spending behavior, yielding measurable benefits involving revenue growth and increased customer retention. However, much of the current research continues to focus mainly on structured booking records, often overlooking unstructured guest reviews that can carry important emotional and experiential details.

In our earlier work, we first developed custom anomaly detection strategies to strengthen the reliability of hotel reviews (Nikolić et al., 2024a) and then introduced traditional data-driven models to identify inconsistent guest feedback (Nikolić et al., 2024b). Building on this foundation, our later work progressed to deep learning techniques for predicting hotel ratings based on observed guest reviews (Nikolić et al., 2025a) and applied transformer-based methods to detect negative feedback with greater accuracy (Nikolić et al., 2025b). Collectively, these efforts achieved strong predictive accuracy (over 90%) and established a foundation for more transparent and trustworthy review analytics in hospitality research.

This paper continues along this path, extending our focus toward a hybrid AI framework that combines structured booking attributes with unstructured guest reviews. The methodology employs RoBERTa embeddings to capture contextual sentiment, while XGBoost algorithm identifies specific patterns in numerical reservation data. By integrating these complementary modalities, the model enhances the accuracy of CLV predictions and offers interpretable results through SHAP analysis, highlighting the importance of key features.

2. PREVIOUS FINDINGS

Recent studies demonstrate the central role of predictive AI in improving hotel management outcomes, especially in forecasting customer behaviors relevant to CLV. Wu and Ma (2025) reported that a hybrid neural network combined with ensemble methods achieved 81.4-82.7% accuracy on 3.2 million records obtained from 846,000 guests, while also improving customer segmentation accuracy by 47.6%. In the same study, they further linked hybrid models to measurable business impacts, including revenue increases of 16.7–23.7%, customer retention gains of 27.3%, loyalty engagement improvements exceeding 42%, and marketing ROI uplifts of 115%. Similarly, Rodrigues et al. (2025) observed recall rates above 80% for predicting cancellations, rebookings, and food package purchases. These findings suggest that intelligent forecasting models can successfully capture various phases of guest engagement across the booking journey.

Likewise, Cheng (2024) applied logistic regression, random forest, and neural networks to predict customer churn, confirming the robustness of classic and modern models put together. Furthermore, Dursun-Cengizci and Caber (2024) achieved 80% churn prediction accuracy using random forest methods on repeat customer datasets. In a large public dataset of 119,386 customers, Choi and Choi (2020) reported an overall loyalty prediction accuracy of 98.9%, with notably higher performance for first-time guests (99.43%) compared to repeat customers (81.79%). Building on these key foundations, Alsharafa et al. (2024) proved that deep neural networks can reduce forecast error to below 12% across multiple evaluation metrics (MAPE, MSE, RMSE), outperforming decision-tree and random-forest baselines.

Related studies provided by Buhalis et al. (2022) and Gatera (2024) underscore the value of integrating forecasting CLV with revenue management workflows, providing smarter pricing, reduced online travel agency commissions, and notably improved revenue per available room (RevPAR). Adhegaonkar et al. (2024) further demonstrated that automated machine learning systems can match the performance of manually optimized models on moderate-sized CLV datasets, thereby simplifying deployment for hotels lacking advanced data science teams.

Despite these advances, most existing models remain focused on structured transactional and behavioral data. While features such as recency, frequency, and monetary value remain highly predictive, they often overlook emotional and experiential cues embedded in online reviews. As Shen (2024) argues, explainable AI is progressively necessary to manage influential guest feedback and gain trust in automated decision-making. The limited integration of unstructured review records into CLV models underscores the need for hybrid frameworks that combine booking data with textual sentiment and experiential signals.

3. METHODOLOGY

The methodological framework for this study consists of a multimodal pipeline that integrates structured reservation data with unstructured guest reviews to capture both behavioral and experiential insights for predicting Customer Lifetime Value (CLV). The process unfolds in five phases: identifying data sources, preprocessing inputs, integrating RoBERTa embeddings with booking patterns extracted using XGBoost, conducting supervised model training with optimization, and evaluating performance through SHAP-based interpretability.

3.1. Data Sources

The empirical analysis relies on a multi-input dataset that combines reservation records with narrative customer feedback, providing a clear representation of behavioral transactions and subjective experiences. The structured component involved operational booking details such as stay duration, advance booking time, frequency of modifications, cancellation history, and the presence of special requests, along with guest composition (adults, children, and babies), followed by average daily rate, repeated indicators for guests, and distribution channels. The unstructured component consisted of reviews written by guests, enriched with numeric scores, reviewer metadata (including guest nationality and trip type), timestamps, and descriptions of features like service quality, cleanliness, staff interactions, and overall satisfaction.

Together, these cooperative data collections supported the integration of quantitative booking behaviors with qualitative impressions, forming a solid foundation for CLV prediction. All reviews were sourced from publicly available datasets, ensuring transparency, reliability, and the possibility of replication in future studies.

Specifically, this analysis uses the *Hotel Booking Demand Dataset*, with 119,390 reservation records across resorts and city hotels in Portugal (Antonio et al., 2019), then the *Booking.com 515K Hotel Reviews Data in Europe*, with approximately 515,000 reviews from global properties (Liu, n.d.), and the *TripAdvisor Hotel Reviews Dataset*, containing approximately 878,500 reviews with ratings and narrative feedback (Arvidsson, n.d.).

To ensure linguistic consistency in the dataset, only English-language reviews were retained. In addition, sentiment variables were derived from the obtained textual data, capturing service features like cleanliness, amenities, distance, and comfort, with each aspect assigned scores based on sentiment. Building on our prior research, inconsistent or anomalous reviews were filtered out to enhance data integrity and reduce noise in the modeling process.

Furthermore, considering that datasets did not share a common hotel identifier, the integration was performed at the feature level rather than directly mapping individual hotels. Structured reservation information from the booking dataset and sentiment-driven attributes from the reviews were combined in a unified fusion layer, enabling the model to learn from behavioral, transactional, and experiential signals without requiring a one-to-one hotel matching.

This integration strategy addresses limitations of prior CLV research, which has relied mainly on transactional records. While these capture frequency and monetary value, they overlook emotional and experiential aspects that shape loyalty. By linking booking patterns with guest narratives, this study offers a more accurate and interpretable measure of CLV that reflects both financial behavior and customer experience.

3.2 Data Preprocessing

To prepare the multimodal dataset for analysis, a series of preprocessing steps was applied. The structured reservation data required standardization and cleaning to maintain consistency across sources. Variables involving length of stay, lead time, cancellation status, and special requests were normalized into common units and formats. Missing values were imputed using median substitution for numerical features, and mode replacement for categorical features, while outliers (e.g., excessively long stays or unrealistic lead times) were discarded to reduce skewness. All categorical features were encoded into numeric form using one-hot encoding to support their direct use in machine learning models. A summary of the datasets and variables retained after preprocessing is provided in Table 2.

Table 2. Dataset statistics before and after preprocessing

Dataset	Raw Records	Preprocessed Records	Key Features
Hotel Booking Demand	119,390	~115,000	Reservation details, guest types, dates, cancellations, modifications, special requests
Booking.com Hotel Reviews	~515,000	~475,000	Removed inconsistent reviews, retained ones with key reviewer metadata and star ratings
TripAdvisor Hotel Reviews	~878,500	~720,000	English-only reviews, aspect-based sentiments (cleanliness, amenities, distance, comfort)

Source: Own processing

Moreover, the unstructured review records required more extensive text preprocessing. Raw text was first cleaned to remove unnecessary punctuation, HTML tags, numerical strings, and special characters. Stopwords were eliminated to reduce noise, and all words were lowercased to maintain uniformity. Lemmatization was then applied to reduce words to their base forms, improving semantic consistency. Once the text was cleaned and normalized, it was tokenized and transformed into dense semantic representations using the RoBERTa model, selected for its strong contextual accuracy and robustness in domain-specific sentiment analysis. These embeddings captured underlying components in guest narratives, especially opinions about cleanliness, staff professionalism, or service quality.

After preprocessing, the structured and unstructured features were consolidated into a unified representation. The representation was used as input for forecasting CLV, with the model outputting predicted CLV scores that distinguish between low-, medium-, and high-value customers. Figure 1 illustrates the overall workflow of these steps.

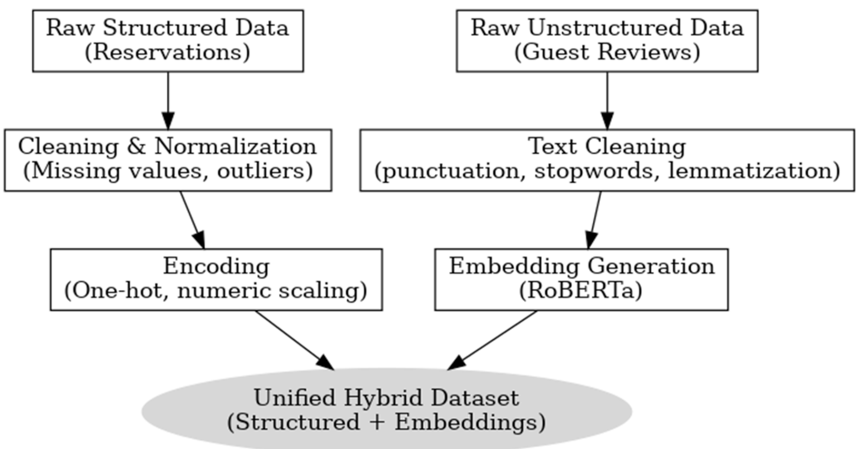

Figure 1. Preprocessing workflow for hybrid dataset construction.
Source: Own processing

3.3. Hybrid Integration

The core methodological innovation of this study lies in advancing predictive CLV modeling through the integration of heterogeneous data sources. As already highlighted in the literature review, a robust CLV prediction requires methods capable of handling complexity, ensuring interpretability, and aligning with strategic decision-making. To address these requirements, XGBoost was utilized because of its efficiency in processing high-dimensional tabular inputs, followed by abilities to model nonlinear interactions, and its resistance to overfitting, making it one of the most widely adopted algorithms for structured data (Adegoke, 2025).

In parallel, RoBERTa embeddings were introduced to represent guest experiences. As a deep transformer-based language model, RoBERTa excels at capturing context-dependent meaning by considering full sentence structure, which allows it to detect subtle shifts in sentiment (e.g., positive versus negative tones in similar phrases) and extract details from service evaluations. Unlike bag-of-words or TF–IDF representations, which consider text as unordered collections of tokens, RoBERTa retains syntactic structure and long-range dependencies between words. This capability is particularly important in hospitality reviews, where meaning frequently depends on nuanced qualifiers (e.g., *"small room but excellent service"*) and multi-aspect evaluations within the same sentence (Ibitoye et al., 2025).

In this context, RoBERTa was selected over alternatives like BERT and DistilBERT because it benefits from extended pretraining and dynamic masking, which typically demonstrate stronger performance on sentiment-based review tasks and other NLP benchmarks. DistilBERT offers greater efficiency but with lower contextual accuracy, which is critical when misclassifying high-value CLV segments carries a disproportionate managerial cost and may lead to inaccurate targeting decisions (Vijay & Premjith, 2024). Domain-specific transformer variants (including FinBERT for financial narratives and SciBERT for scientific texts) were also considered, but these models are optimized for highly specialized vocabularies and discourse patterns that differ from hospitality reviews (Tzimiris et al., 2025). Adapting them to hotel contexts without large in-domain corpora would increase the risk of overfitting and limit generalizability across hotel markets and operational environments. For these reasons, the RoBERTa model clearly provided the optimal balance between predictive accuracy and robustness required for decision-support systems in hotel management.

This fusion was implemented to establish a balanced contribution from both types of data. The resulting joint representation was then passed to the classification layer responsible for predicting CLV categories. Conceptually, the hybrid design follows a two-stream architecture in which input records are first processed independently, optimized within their respective modalities, and subsequently integrated to leverage complementary strengths.

The key rationale behind the hybrid approach is that neither behavioral variables nor textual expressions alone fully capture customer value. Guests with similar spending patterns may express very different intentions in reviews, and conversely, highly positive reviews do not always translate into high monetary value. The hybrid model therefore allows both perspectives to be represented simultaneously within the predictive process.

As noted in the data preprocessing stage, the datasets do not share common hotel identifiers, so integration is carried out at the feature level rather than by linking individual hotels or guests, which notably reduces the risk of data leakage, spurious correlations, or artificially inflated performance.

This structure encourages the hybrid model to learn generalizable relationships between booking behaviours and typical review semantics, rather than exploiting duplicated information about the same customer or property or memorizing specific patterns. This also aligns with best practices in multimodal learning, where the goal is not to perfectly reconstruct a specific dataset, but to extract cross-domain patterns that remain meaningful when applied to new hotels or market contexts.

To provide greater transparency, the hybrid model can be described in more detail as follows. In the structured-data stream, reservation attributes are provided as input to an XGBoost model that learns nonlinear relationships between behavioral indicators such as frequency of stays, length of stay, lead time, cancellations, and special requests. In the textual stream, review texts are first preprocessed and then encoded into 768-dimensional contextual embeddings using RoBERTa. These embeddings summarize semantic information relating to service quality, emotions expressed in reviews, perceived value, and guest satisfaction. In practical terms, this means that the model can account not only for *what* guests do (transactions and bookings) but also for *how they feel and what they report doing* during their stay, integrating emotional tone and perceived service quality as explicit determinants of CLV.

All preprocessing transformations and model-fitting steps are learned on the training partition only and subsequently applied to the validation and test sets, ensuring that no information from the held-out data leaks into the training process. This strict separation of training and evaluation is essential for obtaining unbiased estimates of out-of-sample performance, especially in high-capacity hybrid architectures. It also supports reproducibility, since the same preprocessing and training methods can be applied when the model is deployed on new hotels or updated datasets.

The outputs of the two streams are concatenated in a fusion layer that integrates transactional and experiential information into a unified multimodal feature vector structure. Importantly, the interaction between modalities occurs only at this fusion stage, meaning that neither stream constrains nor dominates the learning dynamics of the other. This strict separation preserves interpretability, as the effect of structured and unstructured predictors can be further inspected independently using feature-importance scores and SHAP analysis.

This design decision was intentional, as early fusion approaches can blur modality-specific effects, whereas late fusion would limit interaction across data types. The chosen intermediate-level fusion strikes a balance by enabling interaction while still allowing each modality to retain its explanatory contribution.

In practice, the fusion layer operates on a fixed-size vector obtained through concatenating the RoBERTa embedding with the latent representation learned for structured features, which is then passed to the classification head.

The fused feature vector is subsequently fed into a fully connected classification layer that outputs the probabilities of belonging to each CLV category (low, medium, high). During training, the two component models are optimized within their own domains, after which parameters of the classification head are updated based on the combined representation. This design allows the architecture to benefit from the expressiveness of deep language models and retain the stability and transparency of gradient-boosted decision trees applied to tabular data.

A detailed block-diagram architecture is provided in Figure 2, illustrating all input streams, the RoBERTa and XGBoost processing blocks, the multimodal fusion layer, and the output nodes.

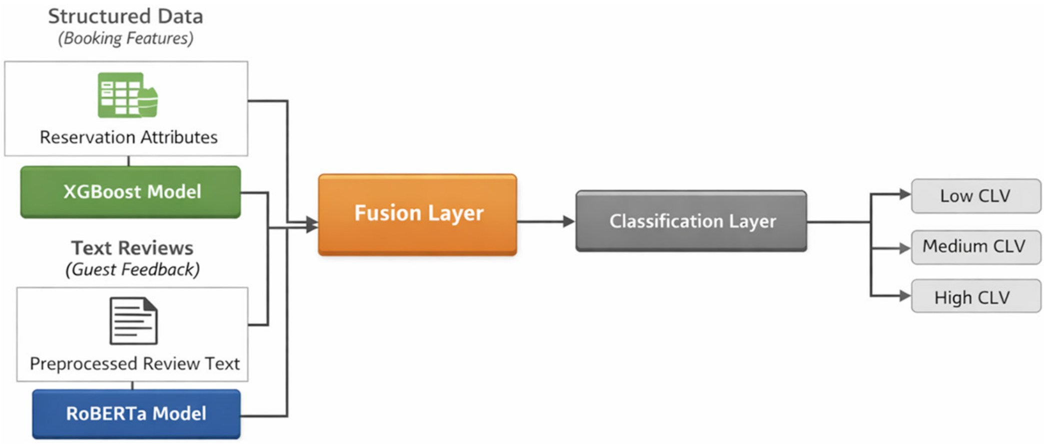

Figure 2. Hybrid CLV architecture combining XGBoost on booking features with RoBERTa review embeddings via a fusion layer and final classification layer

Source: Own processing

For clarity, the overall procedure can be summarized in pseudo-algorithmic form:

```
Input:
Structured booking variables (D_struct)
Raw review texts (D_text)
CLV labels (y)

Output:
Trained hybrid CLV model
Evaluation metrics (accuracy, precision, recall, F1-score)

Step 1: Data preparation
        Extract structured features X_struct from D_struct.
        Extract raw review texts T from D_text.
        Split (X_struct, T, y) into stratified
        training, validation, and test sets.

Step 2: Structured stream
        2.1  Preprocess X_struct (cleaning, encoding, scaling).
        2.2  Train XGBoost component on structured training data.

Step 3: Textual stream
        3.1  Clean review texts (remove noise, normalize).
        3.2  Generate RoBERTa embeddings for each review.
        3.3  Optionally compute sentiment-based features from the cleaned texts.

Step 4: Fusion layer
        4.1  For each sample, concatenate:
             structured representation from XGBoost
             textual representation from RoBERTa (and sentiment features).
        4.2  Obtain a unified multimodal feature vector Z.

Step 5: Classification head
        5.1  Train a fully connected classification layer on Z to predict CLV
             categories (low, medium, high)using the training set.
        5.2  Validate the model on the validation set and tune hyperparameters
             if required.

Step 6: Evaluation
        6.1  Apply the final hybrid model to the test set.
        6.2  Compute accuracy, precision, recall, and F1-score.
        6.3  Report performance and interpret feature contributions using SHAP
             for structured and textual predictors.
```

Taken together, the hybrid architecture operationalizes a comprehensive view of the customer, combining revenue-relevant behavioral attributes with subjective experiential assessments, which is consistent with contemporary conceptualizations of CLV in hospitality management.

As a comparative baseline, we also implemented a simpler fusion model in which sentiment scores derived from guest reviews were concatenated with structured booking features and used as input to an XGBoost classifier, allowing us to evaluate whether the proposed RoBERTa-based hybrid architecture provides value beyond conventional feature-level fusion.

3.4. Model Training and Optimization

The training process was designed to ensure that both data streams were optimized within their respective modalities before being fused into the hybrid framework. For the structured data stream, the XGBoost algorithm was trained with hyperparameter tuning, including the number of estimators, maximum tree depth, learning rate, and subsampling ratio. Additional adjustments such as column sampling and regularization terms were explored to balance bias and variance. Grid search combined with five-fold cross-validation was employed to identify parameter settings that minimized overfitting and maximized predictive accuracy.

Training of the unstructured module included fine-tuning with an Adam optimizer, an initial learning rate of 2e-5, and early stopping criteria based on validation loss. A dropout layer was applied during training to reduce overfitting and enhance the generalization capacity of the embeddings. This optimization strategy was inspired by earlier research demonstrating the effectiveness of low learning rates and dropout regularization in improving stability for text-based models (Mirabdolbaghi & Amiri, 2022).

After the two data streams were individually optimized, their outputs were concatenated in a multi-domain fusion layer. This representation was passed to a classification layer responsible for predicting Customer Lifetime Value (CLV) categories. The entire model was trained in a supervised manner using a stratified train-validation-test split technique, with 70 percent of the data used for training, 15 percent for validation, and 15 percent for testing. Stratification ensured that all CLV categories were proportionally represented in each split.

To further enhance robustness, the hybrid model was trained under multiple random seeds, and the results were averaged to reduce variance. Regularization techniques like L2 penalty were additionally incorporated into the classification layer to effectively mitigate overfitting.

3.5. Evaluation Metrics and Interpretability

The performance of the hybrid approach was evaluated with standard classification metrics: accuracy, precision, recall, and the F1-score. Accuracy provided a general measure of overall correctness, while precision and recall captured the model's ability to distinguish between high, medium, and low-value customers. The F1-score was further obtained as a balanced metric, considering the unequal distribution of CLV categories. To ensure effectiveness, results were reported on the test set and validated through repeated training with multiple random seeds.

In addition to forecasting performance, interpretability was incorporated to ensure practical relevance for hotel management. Using SHAP (SHapley Additive exPlanations), each feature was assigned a numerical contribution to CLV predictions, with the analysis underscoring structured attributes and semantic indicators from reviews. SHAP is useful in this case since it quantifies the relative weight of features. For instance, repeat bookings might have a SHAP score of 0.35 compared to 0.12 for lead time (Meng et al., 2020). This transparency makes it clear how different factors influence the model, as further illustrated in the results section.

3.6. Reproducibility and Computing Environment

To ensure reproducibility and methodological transparency, all experiments were conducted in a specified software and hardware environment. The full pipeline, including preprocessing, embedding generation, model training, and evaluation, was implemented in Python 3.10**.** Structured-data models were developed using scikit-learn and XGBoost, while the text module employed PyTorch along with HuggingFace Transformers for RoBERTa fine-tuning.

Model development was performed on a workstation equipped with an NVIDIA RTX 3060 GPU (12 GB VRAM)**,** Intel i7 processor**,** and 32 GB RAM**.** Random seeds were fixed to 42 across libraries (NumPy, PyTorch, and XGBoost) to guarantee deterministic data splits and initialization procedures. The average RoBERTa fine-tuning run required approximately 2-2.5 hours, whereas XGBoost training completed in under three minutes per configuration.

To avoid risk of data leakage or optimistic bias, all preprocessing transformations, embedding training, and hyperparameter optimization were performed exclusively on the training partition and subsequently applied to validation and test data. Cross-dataset integration occurred strictly at the feature level, without shared identifiers across datasets, making sure that the hybrid model learned generalizable patterns rather than memorizing specific hotels or guests.

Table 3. Final hyperparameters used in model training

Component	Hyperparameter	Value
RoBERTa module	Learning rate	2e-5
	Batch size	16
	Epochs	4
	Max sequence length	256
	Optimizer	AdamW
	Dropout	0.1
	Random seed	42
XGBoost module	Number of estimators	300
	Max depth	6
	Learning rate	0.05
	Subsample	0.8
	Column subsample (colsample_bytree)	0.8
	L2 regularization (lambda)	1.0
	Random seed	42

Source: Own processing

Because CLV classes were imbalanced in the underlying data, experiments were repeated with class-weighted loss functions and stratified sampling, and results were averaged across multiple random seeds to ensure consistency of the estimates. Oversampling techniques such as SMOTE were deliberately not applied to avoid introducing synthetic samples that could distort the joint distribution between reservation patterns and review semantics. To further control variance and avoid overfitting, hyperparameters for both components were selected through five-fold cross-validation on the training data, and the final values from experiments are presented in Table 3.

4. EXPERIMENTAL RESULTS

This section presents the empirical evaluation of the proposed hybrid framework. First, we report quantitative classification results obtained across the CLV categories and compare the hybrid model with alternative learning strategies. Performance is assessed through accuracy, precision, recall, and F1-score. Next, the managerial and practical relevance of the model is discussed through concrete application scenarios that show how predicted CLV segments can inform hotel revenue management, personalization strategies, and customer retention policies.

4.1. Predictive Performance

As a first step, a simpler fusion baseline was evaluated in which sentiment scores extracted from guest reviews were concatenated with structured booking features and classified using XGBoost. This baseline already improved performance compared with single-source models, reaching approximately 82% accuracy, 80% precision, and 79% recall**,** indicating that even basic sentiment-enhanced fusion can capture additional information beyond purely numerical reservations or text alone.

The evaluation of the proposed hybrid framework, however, demonstrates a further and more substantial improvement in predictive performance. As shown in Figure 3, the model that combines RoBERTa embeddings of guest reviews with XGBoost on structured reservation features consistently outperforms models trained on reservations or reviews only. The hybrid approach achieved approximately 89% accuracy, 87% precision, and 86% recall, whereas the model based on reviews only reached around 72% across metrics and the model based on reservations only remained close to 63%. The radar chart below highlights how the RoBERTa with XGBoost method maintains a balanced advantage in all crucial measures of estimation quality, confirming that behavioral signals from numerical reservation data and experiential cues from guest reviews are complementary in forecasting Customer Lifetime Value.

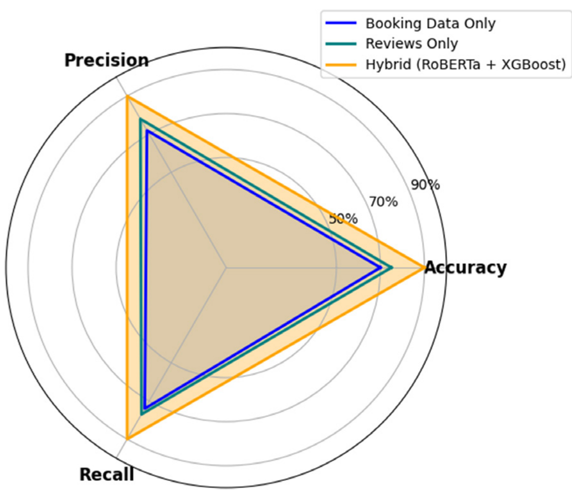

Figure 3. Performance comparison of three models across key metrics
Source: Own processing

This reduction of over 20% compared to single-source models has important implications for hotel management, since misclassifying high-value clients can result in missed opportunities for personalization, weakened loyalty programs, and inefficient resource allocation.

Beyond overall accuracy, the analysis also considered how the model manages class imbalances among low, medium, and high CLV categories. Confusion matrices and classification reports indicated that the hybrid method notably reduced the number of false negatives, particularly cases where high-value customers were misclassified as lower-value segments. Compared with the sentiment-fusion baseline, the hybrid model also achieved higher recall for the high-CLV segment, suggesting that contextual RoBERTa embeddings provide discriminative information that simple polarity scores cannot fully capture.

Interpretability was also an important component of the evaluation, as predictive models must provide transparency into the factors that influence their outcomes. SHAP (SHapley Additive exPlanations) analysis was employed to quantify feature contributions, revealing that both experiential and behavioral indicators are significant. Reviews emphasizing cleanliness, staff professionalism, and service quality were recognized as the strongest predictors of high CLV. At the same time, structured attributes like repeat bookings and special requests contributed substantially, refining the classification of guests likely to build relationships with a hotel. The results are given in Figure 4, showing the importance of structured and unstructured predictors.

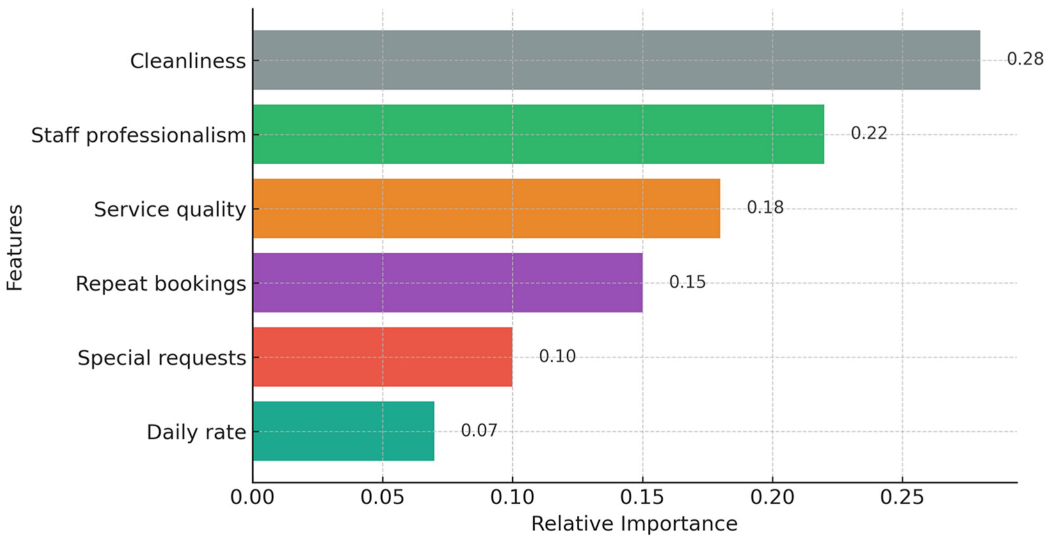

Figure 4. Relative importance of structured and unstructured features in predicting CLV
Source: Own processing

4.2. Managerial Applications

Beyond numerical improvements in classification metrics, the model has direct implications for hotel decision-making. Predicted CLV segments provide a basis for differentiated customer treatment in several operational domains. In practice, this turns CLV labels into concrete service rules. High-CLV guests identified by the hybrid model can be prioritized for proactive service recovery, complimentary upgrades, and personalized communication, increasing the likelihood of long-term retention and advocacy. Medium-CLV customers may be targeted through tailored loyalty incentives or cross-selling strategies designed to stimulate repeat visits and gradual value escalation. Low-CLV guests can be managed using cost-efficient service configurations, helping hotels to allocate resources without compromising perceived service quality.

The hybrid model also supports data-driven segmentation strategies. By jointly analyzing transactional histories and experiential signals in reviews, hotels can distinguish, for example, between frequent but price-sensitive guests and less frequent yet high-spending customers with strong satisfaction indicators. In practice, these segment differences can be translated into clear campaign rules. These distinctions provide precise campaign design in customer relationship management systems, including personalized email marketing, pricing differentiation, and upselling of auxiliary services such as late checkout or dining packages.

Finally, the model provides early-warning capacity. Negative textual sentiment combined with previously high CLV scores may indicate guests at risk of churn, allowing hotels to intervene through tailored outreach. In this way, the framework not only classifies customers but also supports retention planning, revenue optimization, and long-term relationship management.

5. CONCLUSION

This research proposed and validated a hybrid predictive framework for forecasting Customer Lifetime Value (CLV) in the hospitality industry by aligning structured reservation data with unstructured guest reviews. By integrating behavioral and transactional indicators developed through XGBoost with contextual embeddings generated from RoBERTa, the methodology captured objective booking patterns and subjective experiential feedback. The multi-source fusion achieved higher forecasting accuracy (~89%) than models relying on either data source alone (63% and 72%), demonstrating the value of linking cooperative information streams.

The results advance both technical methodology and practical management. From a technical perspective, the results highlight how multimodal architectures can improve model robustness and interpretability, with SHAP analysis revealing the relative importance of features such as repeat bookings, length of stay, cleanliness, and service quality. From a business perspective, the proposed model enables hotels to effectively identify high-value customers and tailor engagement strategies that enhance retention and profitability.

By reframing CLV prediction as a multimodal AI task, this study shows that hotels can move from intuition-driven decisions toward systematic and evidence-based management of guest relationships, where resource allocation, personalization, and retention efforts are efficiently guided by quantified insight rather than informal judgment.

The importance of this study also lies in bridging the gap between advanced AI methods and practical decision-making in hotel management, offering a scalable and interpretable solution that can be seamlessly integrated into existing customer relationship management systems. By supporting more precise targeting and refined personalization, such strategies help hospitality providers to strengthen customer loyalty, lower churn, and maximize revenue streams. In the longer term, the adoption of hybrid AI models may fundamentally transform how hotels manage guest relationships.

Despite these contributions, the study has several limitations. The hybrid model is trained on three public datasets from specific countries and online platforms, so the learned patterns may not fully generalize to individual properties or markets with different demand structures and review cultures. In addition, the evaluation relies on offline historical data rather than live deployment, meaning that the true business impact on revenue, loyalty, and churn still needs to be confirmed through field experiments or A/B testing.

Future research could extend this work by incorporating temporal dynamics of customer behavior, experimenting with alternate fusion strategies (e.g., attention-based mechanisms or graph-based fusion), and validating the model across diverse hospitality settings to ensure generalizability and resilience.

Acknowledgment

This research was supported by the Science Fund of the Republic of Serbia, Grant No. 7502, Intelligent Multi-Agent Control and Optimization applied to Green Buildings and Environmental Monitoring Drone Swarms – ECOSwarm.

References

Adegoke, A. (2025). Leveraging Predictive Customer Lifetime Value in Financial Consulting to Drive Growth-Focused Service Redesign Strategies. *International Journal of Research Publication and Reviews, 6*(6), 189-211. https://doi.org/10.55248/gengpi.6.0625.2220

Adhegaonkar, V. R., Thakur, A. R., Varghese, N., & Cheriyan, A. (2024, March). Decoding Customer Lifetime Value to Unlock Business Success with Predictive Machine Learning Approach. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies* (pp. 1-4). IEEE. https://doi.org/10.1109/tqcebt59414.2024.10545084

Alsharafa, N. S., Madhubala, P., Moorthygari, S. L., Rajapraveen, K. N., Kumar, B. R., Sengan, S., & Dadheech, P. (2024). Deep learning techniques for predicting the customer lifetime value to improve customer relationship management. *Journal of Autonomous Intelligence*, *7*(5). https://doi.org/10.32629/jai.v7i5.1622

Antonio, N., de Almeida, A., & Nunes, L. (2019). Hotel booking demand datasets. *Data in Brief, 22*, 41–49. https://doi.org/10.1016/j.dib.2018.11.126

Arvidsson, J. (n.d.). *TripAdvisor hotel reviews* [Dataset]. Kaggle. https://www.kaggle.com/datasets/joebeachcapital/hotel-reviews

Buhalis, D., O'Connor, P., & Leung, R. (2022). Smart hospitality: From smart cities and smart tourism towards agile business ecosystems in networked destinations. *International Journal of Contemporary Hospitality Management, 35*(1), 369–393. https://doi.org/10.1108/ijchm-04-2022-0497

Cheng, J. (2024). AI-Based Hotel Customer Churn Prediction Model. *Journal of Progress in Engineering and Physical Science*, *3*(4), 15–21. https://doi.org/10.56397/jpeps.2024.12.03

Choi, Y. O., & Choi, J. (2020). The prediction of hotel customer loyalty using machine learning technique. *International Journal of Advanced Trends in Computer Science and Engineering, 9*(5), 7395–7402. https://doi.org/10.30534/ijatcse/2020/143952020

Dursun-Cengizci, A., & Caber, M. (2024). Using machine learning methods to predict future churners: An analysis of repeat hotel customers. *International Journal of Contemporary Hospitality Management, 37*(1), 36–56. https://doi.org/10.1108/ijchm-06-2023-0844

Gatera, A. (2024). Role of Artificial Intelligence in Revenue Management and Pricing Strategies in Hotels. *Journal of Modern Hospitality*, *3*(2), 14–25. https://doi.org/10.47941/jmh.1957

Henriques, H., & Pereira, L. N. (2024). Hotel demand forecasting models and methods using artificial intelligence: A systematic literature review. *Tourism & Management Studies*, *20*(3), 39-51. https://doi.org/10.18089/tms.20240304

Ibitoye, A. O. J., Kolade, O., & Onifade, O. F. W. (2025). Customer retention model using machine learning for improved user-centric quality of experience through personalised quality of service. *Journal of Business Analytics,* 1–19. https://doi.org/10.1080/2573234X.2025.2551950

Kabir, F., Khan, M. R., Mia, M. N., & Talukder, M. B. (2024). Implications of Artificial Intelligence (AI) in the Hotel Industry. In *Hotel and Travel Management in the AI Era* (pp. 357-378). IGI Global. https://doi.org/10.4018/979-8-3693-7898-4.ch017

Liu, J. (n.d.). *515K hotel reviews data in Europe* [Dataset]. Kaggle. https://www.kaggle.com/datasets/jiashenliu/515k-hotel-reviews-data-in-europe

Meng, Y., Yang, N., Qian, Z., & Zhang, G. (2020). What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(3), 466-490. https://doi.org/10.3390/jtaer16030029

Mirabdolbaghi, S. M. S., & Amiri, B. (2022). Model optimization analysis of customer churn prediction using machine learning algorithms with focus on feature reductions. *Discrete Dynamics in Nature and Society, 2022,* Article 5134356. https://doi.org/10.1155/2022/5134356

Nikolić, M., Stojanović, M., & Marjanović, M. (2024a). Anomaly detection in hotel reviews: Applying data science for enhanced review integrity. *2024 32nd Telecommunications Forum (TELFOR),* 1–4. https://doi.org/10.1109/TELFOR63250.2024.10819036

Nikolić, M., Stojanović, M., & Marjanović, M. (2024b). Integrating data science and predictive modeling for detecting inconsistent hotel reviews. *International Scientific Conference UNITECH 2024 – Selected Papers,* 1–16. Technical University of Gabrovo. https://doi.org/10.70456/DHXA1258

Nikolić, M., Stojanović, M., & Marjanović, M. (2025a). Integrating deep learning for automated detection of negative hotel reviews. *Facta Universitatis, Series: Automatic Control and Robotics,* 1–16. https://doi.org/10.22190/FUACR241218002N

Nikolić, M., Stojanović, M., & Marjanović, M. (2025b). The power of words: Leveraging deep learning techniques to predict hotel ratings from user reviews. *2025 24th International Symposium INFOTEH-JAHORINA (INFOTEH),* 1–6. https://doi.org/10.1109/INFOTEH64129.2025.10959201

Rodrigues, D., Jardim, B., & De Castro Neto, M. (2025). Predicting key touchpoints in hotel customer journey – a comparison of machine learning models. *Journal of Travel & Tourism Marketing, 42*(5), 609–626. https://doi.org/10.1080/10548408.2025.2456083

Shen, R. (2024). From prediction to explanation: Managing influential negative reviews through explainable AI. *arXiv.* https://arxiv.org/pdf/2412.19692v1

Tzimiris, S., Nikiforos, S., Nikiforos, M. N., Mouratidis, D., & Kermanidis, K. L. (2025). A Comparative Evaluation of Transformer-Based Language Models for Topic-Based Sentiment Analysis. *Electronics*, *14*(15), 2957. https://doi.org/10.3390/electronics14152957

Vijay, V. H., & Premjith, B. (2024). Enhancing Financial Sentiment Analysis with Pre-trained BERT-Based Models. In *Congress on Smart Computing Technologies* (pp. 79-90). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-96-6254-8_6

Webb, T., Cho, R., & Legg, M. P. (2022). Customer lifetime value: A data science approach for hospitality applications. *International Journal of Gaming, Hospitality and Tourism, 2*(1). https://www.ijght.org/index.php/light/article/view/46

Wu, D., & Ma, Q. (2025). AI-assisted customer behavior analysis and hotel loyalty strategy optimization. *Future Technology, 4*(3), 97–106. https://doi.org/10.55670/fpll.futech.4.3.10